### 1. **<u>Why The Simple Question May Give A Misleading Answer</u>**

50. A simple example will show the possible difference in these two questions. Consider the variable, mean time for the incumbent LEC to restore service. Suppose that for 100 observations, the incumbent LEC's customers faced a mean time of 3.4 hours with a variance of 0.0145, while the competing LEC's customers had a mean time of 5.0 hours with a variance of 0.0404, also for 100 observations. A simple measure of the difference in this case would reveal a difference of 1.6 hours and a z-statistic which is large enough to imply ample evidence of discrimination.

51. Suppose, however, that the time to restore service is always exactly 3 hours if service is disrupted during the week and is always exactly 7 hours if service is disrupted on a weekend. Further, suppose that 90% of the incumbent LEC's customers had service disrupted on a weekday, but only 50% of the competitor LEC's customers had service disrupted on a weekday.

52. Given 3 hours to restore weekday service versus 7 hours to restore weekend service, we would expect the incumbent LEC's customers to have service restored in 3.4 hours on average ($3 \cdot 0.9 + 7 \cdot 0.1$), whereas we would expect the competitive LEC's customers to have service restored in 5.0 hours on average ($3 \cdot 0.5 + 7 \cdot 0.5$). In other words, the difference in this case can be entirely explained as being the result of differences in the days on which service repair calls were received, and not the result of any discrimination on the part of the incumbent LEC.

### 2. Appropriate Techniques For Identifying Discrimination

53. If the purpose of this analysis is to identify instances of discrimination by the incumbent LEC, then simply measuring the significance of differences in variables is not the appropriate way of identifying discrimination.

54. Given the necessary data and a reasonable amount of time, the most appropriate technique for answering the question, Why is there a significant difference? is to determine what factors affect the variable (e.g., response time), and to attempt to explain the variable as a function of those variables which may affect the variable (e.g., day of week service is disrupted, number of other customers also affected, weather, etc.). If one of the factors that is found to affect the variable in a significant way is whether the customer is a customer of the incumbent LEC or a competing LEC, then there is strong statistical evidence that the incumbent LEC is discriminating against the customers of the competing LEC.

There are a variety of ways in which the question of why there is a significant difference can be addressed. Several of these are discussed below.

#### a. Subdividing Variable into Multiple Groups

55. If there are certain factors which are known (or believed) to affect the variables being evaluated, then a simple way to control for changes in these variables is to sub-divide the variable of interest into multiple groups which control for these other factors.

56. For example, suppose that the time for service to be restored varies by the day of the week. Rather than simply comparing the mean for all customers of the incumbent LEC with all customers of the competing LEC, customers could be divided into fourteen

groups: customers of the incumbent LEC who called on Monday, incumbent customers who called on Tuesday, ..., competing customers who called on Monday, ....

57. Then, if the mean for incumbent LEC customers who called on Monday is compared to the mean for competing customers who called on Monday (and likewise for the other days), this will measure the significance of differences, controlling for differences across days of the week.

58. The problem with this technique is that it requires enough data that, when the data is divided into smaller samples, each of these smaller samples still has enough data to perform the statistical tests described earlier. This is an important limitation to this technique, especially if there is more than one additional factor which may affect a variable (e.g., sub-dividing by day of the week as well as by location). It is, however, possible to control for other factors without these data limitations through the use of multivariate regression analysis.

### b. <u>Multivariate Regression Analysis</u>

59. As an alternative to sub-dividing the samples into smaller groups, one could simply estimate an equation to explain the performance measure as a function of those variables which may affect the performance measure.

60. For example, the time for service to be restored could be specified as a function of the day of the week on which service is disrupted, the number of other customers affected, the weather, and other factors. To investigate the specific question of whether there is any discrimination, one would then include one additional variable as a possible [redacted]

or a competing LEC. For example, the time for service to be restored would be specified as follows:

$$\text{Time} = \text{F(day, COS, etc.)} + \text{DLEC}$$

where F(day, COS, etc.) captures the effect of other factors, and DLEC is a dummy variable equal to zero if the customer is a customer of the incumbent LEC and equal to one if the customer is a customer of a competing LEC.

61. If the coefficient of DLEC in the resulting equation is positive, then, all other things being equal, the time to restore service is longer for customers of competing LECs than for customers of the incumbent LEC. A simple t-test of the coefficient on DLEC can then be used to gauge whether this difference is significant, holding all other factors constant.

62. Multivariate regression analysis along these lines is a powerful tool for evaluating the presence and significance of discrimination. It is much more time-consuming to accurately specify what factors affect each of 30-40 variables than to calculate simple ratios for each of these variables, however. In addition, the specifications chosen in this way may be controversial, with the inclusion of one or more variables affecting the ultimate results.

### c. <u>**Discriminant Analysis**</u>

63. The purpose of discriminant analysis is to develop an equation to predict whether a customer is a customer of the incumbent LEC or of a competing LEC based on observing the variables maintained by the Commission. If these variables are significant in helping

to predict whether a customer belongs to the LEC or the CLEC then this may be interpreted as evidence of discrimination.

64. In a certain respect, discriminant analysis is superior to multivariate regression analysis as an objective statistical tool insofar as the set of independent variables is not subject to judgment. Discriminant analysis requires fairly significant assumptions, however. The key assumptions for discriminant analysis are multivariate normality of the independent variables and equal dispersion and covariance structures for the groups as defined. In addition, discriminant analysis is not useful in directly comparing means of groups. Hence, it is not clear how discriminant analysis would fit into an overall evaluation process based on comparing means of two groups over a set of variables.

## IV. Specific Areas of Comment Requested by the Commission

The Commission raises several additional issues which warrant some further discussion.

### a. One-Tailed versus Two-Tailed Test

65. As the Commission notes, in this case, differences in the variables are only indicative of discrimination if the difference is in one direction (presumably the Commission is not concerned about possible discrimination by LECs against their own customers). Hence, a one-tailed test would appear to be most appropriate for evaluating discrimination.

### b. Desirable Sample Size for Calculations

66. The Commission questions whether small sample sizes, particularly samples of fewer than 30 observations, might render all or some of these proposed statistical tests invalid. This is a legitimate concern.

67. The advantages of a large sample size (typically a sample size of roughly 30 or greater is considered "large" by statisticians) in statistical testing are two-fold. First, large samples enable one to assume that the sample mean is normally distributed, as opposed to being distributed as a Student's t-distribution. The normal distribution is more standardized and is more familiar to most people (e.g., the normal distribution has the familiar result that the 95% confidence level is approximately equivalent to two standard deviations beyond the mean). Fundamentally, however, the Student's t-distribution works equally well as a statistical measure for sample sizes less than 30.

68. The more significant advantage of having large sample sizes is the Central Limit Theorem, which says that for a large enough sample, the mean of any variable with finite variance is approximately normal. This means that for large sample sets, the assumption underlying much of this work that the variables are normally distributed will be at least approximately true.

69. For variables with smaller samples, it will be more important to determine whether the sample distributions are still sufficiently normal to permit the use of a Student's t-distribution. If the distributions of these samples are not normal and cannot be made normal by simple transformation (e.g., a log-normal distribution, in which the natural logarithm of the data are normally distributed), then the statistical tests proposed here may provide unreliable results.

70. There is a statistical basis for determining what constitutes a large enough sample size depending on how precise an estimate is desired. If the population variance is unknown,

as is virtually always the case, then the optimal minimum sample size can be estimated by the following formula:

$$n = [z^2 / H^2] \bullet s^2$$ [6]

where z is the critical value of the t-statistic (e.g., z=1.96 for a 2-tailed 95% confidence interval using the normal distribution), H is the largest absolute deviation we are willing to tolerate, and $s^2$ is the sample variance.

71. For example, suppose that it is determined that a difference in mean time to restore service of more than 0.5 days is viewed as too extensive. Further, assume that the critical value used in evaluating the significance of this difference is 1.645 (a 5% one-tailed test using a normal distribution). Finally, suppose that the sample variance in time to restore service is found to be equal to 1 day. In this case, n would be equal to $[(1.645)^2 / (0.5)^2] \bullet 1 = 10.8$. Hence, in this case, a sample size of 10 observations would yield the desired level of precision.

72. If the number of observations available is less than the optimal value of n, then statistical estimation may be problematic. One possible means of addressing a shortcoming in the amount of data available could be to estimate problematic data on a less frequent basis. For example, if n = 10 and only 5 data points are available in any typical month, there will still be enough data to calculate statistical significance on, in this example, a quarterly basis (5 data points per month = 15 data points per quarter).

---

[6] Gilbert A. Churchill, Jr., Market Research: Methodological Foundations, 6th Edition (The Dryden Press, New York), 1995, p. 631.

The actual number of observations available for the data to be analyzed can be compared with optimal sample size calculations using the above formula for each of the variables which the Commission seeks to analyze. A comparison of these data may prove helpful in determining how often such analyses should be performed. If, for example, the optimal number of observations can be provided quarterly for most data under consideration, then it would be appropriate to make the calculations discussed here on a quarterly basis.

c. **Decision Rule**

73. AT&T proposes an essentially simple decision rule: no more than 5% of comparisons should fail. This is straightforward and is relatively easy to calculate. It is not, however, the same as a decision rule that the probability of failing, given no discrimination, is 5%.

74. Under AT&T's rule, no more than 1 in 400 (0.25%) tests can fail in two consecutive periods without being judged as evidence of discrimination. As discussed above, the appropriate test should be established so that the number of "extreme" values needed to trigger a failure of the overall parity test would only occur 5 percent of the time when underlying parity actually exists.

75. If one conducts individual tests in each of two periods, the total number of failures will follow a binomial distribution. From the binomial distribution, one can then calculate the number of failures, x, such that the probability of failing x or fewer times is 5%. When calculated in this way, over two time periods, there is a 5% chance that 3 of every 400 tests will actually fail in two consecutive periods when there is no discrimination.

A third way of judging discrimination across time periods would be to jointly test the equality of the means in each time period. If the data are independent across time periods, then this will involve a simple F-test. This test will be somewhat more restrictive, however, so that, in fact, there is a 5% chance that as many as 27 of every 400 tests will fail.

### d. <u>Comments on AT&T Proposal</u>

76. AT&T has submitted a proposal with respect to differences in means, for which the Commission asked for comments. In addition, the Commission has asked whether AT&T and BellSouth's proposals would be appropriate for tests of equality of variances and equality of proportions (Boolean variables).

77. AT&T proposed three criteria to determine incumbent LEC compliance, each of which is discussed briefly below.

78. First, AT&T suggests that performance would be considered nondiscriminatory if no more than 5 percent of comparisons fell outside of a 95 percent confidence interval. As discussed above, the appropriate test should be established so that the number of "extreme" values needed to trigger a failure of the overall parity test would only occur 5 percent of the time when underlying parity actually existed.

79. Next, AT&T recommends that no more than 0.25% of measurements should fail this test in two or more consecutive months. Here, AT&T's rules are based on a simple dichotomy of possible outcomes: LECs either pass a test or they fail a test. Given this way of viewing things, AT&T's proposal is reasonable.

80. Again, as discussed above, the test should be established so that the number of extreme values needed to trigger a failure of the overall parity test would occur 5 percent of the time when underlying parity actually existed. Even if discrimination is not statistically significant in a single period, persistent differences may still be an indication of possible discrimination, particularly if they persist over time.

81. A simple example may illuminate this issue. Suppose a test is run for three consecutive months, and a particular LEC has differences with z-statistics equal to 1.2, 1.3, and 1.2. If the critical value for this test is 1.645 (95% one-tail test), then this LEC will "pass" the test for all three months. Taken individually, this is reasonable, insofar as the probability of a z-statistic equal to 1.2 is 11.5% in the absence of discrimination, while the probability of a z-statistic equal to 1.3 is nearly 10% (9.68% to be exact).

82. Taken together, however, three consecutive the statistics in excess of 1.2 are highly unlikely. In fact, the probability of three consecutive such the statistics (assuming the tests are independent) is 0.15% ($0.115^3$). Viewed in this way, therefore, this seems to be compelling evidence of an indication of possible discrimination.

83. Even if one wished to view the situation as AT&T does, as correctly noted by the Commission, this test is valid only if one assumes that measurements of a particular variable by month are independent. A failure in this particular test may therefore be the result of dependence across months rather than the result of discrimination.

84. A simple test of independence across months is to test for the presence of autocorrelation. If a time series is autocorrelated over time, this means that the value of the variable in this month is a function of the value from the previous month. The simplest test for

autocorrelation is to run a simple regression of the variable on the lag of the variable and test the significance of the resulting coefficient.

85. If the variables being tested are autocorrelated, this would mean that variables that are unusually high in one period will be likely to be unusually high in the next period. For example, suppose that in one month the incumbent LEC's customers had to wait an average of 1 hour for service to be activated while competing LEC's customers had to wait an average of 2.5 hours for the same thing. If the average wait time in the next month were in part a function of this wait time, it would not be too surprising to see this difference persist. In such a case, however, the percentage of tests which fail in two consecutive months may be quite a bit higher than 0.25% and, in fact, may be close to the number of tests which fail in any one month, depending on how strong the autocorrelation is.

86. If desired, a test of autocorrelation across time can be combined with a more formal model to explain the variable of interest as described in Step 2 above.

87. Finally, AT&T proposed that a single difference greater than three standard deviations from zero be treated as evidence of discrimination. If the true difference were normally distributed with a mean of zero, the probability of the sample difference being more than three standard deviations greater than zero (focusing on a one-tailed test) would be 0.13%. Hence, such a result would be extremely unlikely (although certainly not impossible).

88. The differences that will be measured here, particularly if sample sizes are small, will not be expected to follow a normal distribution, but will be expected to follow a

t-distribution, which will be dependent on the number of observations. Suppose, for example, there were fewer than thirteen degrees of freedom for a particular variable. In such a case, the probability of the sample difference being more than three standard deviations greater than zero would be 0.5% or more. While this is still an extremely unlikely event, it is, nevertheless, an event that may occur even in the absence of any discrimination.

89. The use of an arbitrary threshold, above which any difference is automatically attributed to discrimination, is at odds with fundamental statistical theory, which allows for the possibility, no matter how small, of extreme cases. Certainly, this third criterion does not appear to add significantly to AT&T's first two proposals, which should adequately flag any possible discrimination.

### e. <u>Comments on BellSouth Proposals</u>

90. BellSouth has made two proposals. The first is essentially identical to AT&T's proposal.

91. BellSouth's second proposal is that if the difference in the means is positive for three consecutive months that this be viewed as an indication of possible discrimination.

92. If the true mean for the incumbent LEC and the competing LEC are equal, then the probability that the sample mean for the incumbent LEC is greater (or less, whichever implies discrimination) than the sample mean for the competing LEC is approximately 50% (0.5). If the difference in means is independent across months, then the probability that the sample mean for the incumbent LEC is greater than the sample mean for the competing LEC for three consecutive months is equal to $0.5^3$, or 0.125 (12.5%). Hence,

under BellSouth's proposal, there is a 12.5 percent chance that evidence of discrimination will be found where no discrimination exists.

### f. Appropriateness of AT&T and BellSouth Proposals in Testing Variances and Boolean Variables

93. In general, the tests proposed by AT&T and BellSouth would be equally appropriate if applied to differences in variances and differences in Boolean variables. In the case of AT&T's third proposal as well as both of BellSouth's proposals, however, this simply means that these tests are equally invalid for the same reasons as outlined above. In addition, the cautions raised concerning AT&T's second proposal are also equally valid in this case.

# April 1998 Call Completion Summary Report

| ILLINOIS | | Call Attempts | Adjusted Blocked Calls | Call Completion |
|---|---|---|---|---|
| **InterLATA** | | | | |
| | *Ameritech* | 94,636,626 | 47,230 | 100.0% |
| | *TC* | 500,558 | 57 | 100.0% |
| **IntraLATA** | | | | |
| | *Ameritech* | 40,208,205 | 33,388 | 99.9% |
| | *TC* | 7,404,625 | 21,924 | 99.7% |

# April 1998 Call Completion Summary Report

| MICHIGAN | | Call Attempts | Adjusted Blocked Calls | Call Completion |
|---|---|---|---|---|
| **InterLATA** | | | | |
| | *Ameritech* | 7,374,828 | 20,332 | 99.7% |
| | *TC* | 500,603 | 155 | 100.0% |
| **IntraLATA** | | | | |
| | *Ameritech* | 22,609,896 | 68,551 | 99.7% |
| | *TC* | 5,781,254 | 44,998 | 99.2% |

# April 1998 Illinois Interlata

| | Call Attempts | Blocked Calls | Reroute Attempts | Reroute Successes | Adjusted Blocked Calls | Call Completion % | Blocking % |
|---|---|---|---|---|---|---|---|
| AMERITECH | [redacted] | | | | | 100.0% | 0.0% |
| CARRIER 1 | [redacted] | | | | | 100.0% | 0.0% |
| CARRIER 2 | [redacted] | | | | | 100.0% | 0.0% |
| CARRIER 3 | [redacted] | | | | | 100.0% | 0.0% |
| CARRIER 4 | [redacted] | | | | | 100.0% | 0.0% |

# April 1998 Illinois Intralata

| | Call Attempts | Blocked Calls | Reroute Attempts | Reroute Successes | Adjusted Blocked Calls | Call Completion % | Blocking % |
|---|---|---|---|---|---|---|---|
| AMERITECH | | | | | | 99.9% | 0.1% |
| CARRIER 1 | [redacted] | | | | | 100.0% | 0.0% |
| CARRIER 2 | [redacted] | | | | | 99.9% | 0.1% |
| CARRIER 3 | [redacted] | | | | | 100.0% | 0.0% |
| CARRIER 4 | [redacted] | | | | | 100.0% | 0.0% |
| CARRIER 5 | [redacted] | | | | | 100.0% | 0.0% |
| CARRIER 6 | [redacted] | | | | | 99.3% | 0.7% |

# *April 1998 Michigan Intralata*

| | Call Attempts | Blocked Calls | Reroute Attempts | Reroute Successes | Adjusted Blocked Calls | Call Completion % | Blocking % |
|---|---|---|---|---|---|---|---|
| AMERITECH | | | | | | 99.7% | 0.3% |
| CARRIER 1 | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | 99.9% | 0.1% |
| CARRIER 2 | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | 99.9% | 0.1% |
| CARRIER 3 | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | 99.9% | 0.1% |
| CARRIER 4 | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | 100.0% | 0.0% |
| CARRIER 5 | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | 100.0% | 0.0% |
| CARRIER 6 | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | 99.7% | 0.3% |
| CARRIER 7 | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | 97.7% | 2.3% |

# *April 1998 Michigan Interlata*

| | Call Attempts | Blocked Calls | Reroute Attempts | Reroute Successes | Adjusted Blocked Calls | Call Completion % | Blocking % |
|---|---|---|---|---|---|---|---|
| AMERITECH | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | 99.7% | 0.3% |
| CARRIER 1 | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | 99.9% | 0.1% |
| CARRIER 2 | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | 100.0% | 0.0% |
| CARRIER 3 | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | 100.0% | 0.0% |
| CARRIER 4 | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | 99.9% | 0.1% |
| CARRIER 5 | [redacted] | [redacted] | [redacted] | [redacted] | [redacted] | 100.0% | 0.0% |

APPENDIX D

# WHITE PAPER ON PERFORMANCE PARITY

## WITH COMMENTS ON FCC'S NPRM 98-72

Dr. Charles D. Pack †

Dr. James Gordon ‡

Dr. Albert A. Fredericks †

*May 28, 1998*

† Monmouth University, West Long Branch, NJ 07764
‡ CAPE Consulting, Rumson, NJ 07760

# WHITE PAPER ON PERFORMANCE PARITY

### WITH COMMENTS ON FCC'S NPRM 98-72

## INTRODUCTION & CREDENTIALS

Ameritech plans to file with the Commission a proposal for measurement of network performance concerning interconnection with Competing Local Exchange Carriers (CLECs). Ameritech proposes to present credible, on-going evidence that they are providing CLECs in their Region IntraLATA and InterLATA access services that are comparable to what they provide themselves. The FCC's NPRM 98-72, specifically items 96-101 under "a. Trunk Blockage", contains various statements, for comment by Ameritech and others, regarding how such performance is or might be measured.

Ameritech has requested that Monmouth University and CAPE Consulting provide them with a conceptual analysis of various approaches, including the Commission's, with regard to definitions of parity, associated measurements and statistics, where in the network such measurements apply and how the Ameritech results may be developed in a meaningful fashion. That conceptual analysis is provided in this document.

The remainder of this paper provides the authors' views on the above measurement and parity issues. At this point in time, these views may be characterized as a conceptual analysis of parity determination, drawing on the authors' experience with network performance in other contexts. Determination of parity in the present context raises new questions, on which the authors have provided 'expert opinion'. Although the authors believe that the conceptual approach advanced in this paper is feasible and can be implemented in practice, it is emphasized that the approach remains to be further analyzed and validated through detailed study of network data. Analysis of actual network data is necessary in order to refine and develop a workable framework for parity determination. The authors' credentials as expert witnesses follow.

**Charles D. Pack,** BSEE, MS, D.Eng.Sci.
Visiting Professor, Monmouth University, West Long Branch, New Jersey

Charles Pack has a Doctor of Engineering Science degree in Operations Research/IE from Columbia University (1972), an MS in Operations Research from Johns Hopkins University (1967) and a BS in EE from the University of Delaware (1965). He has more than 33 years of experience in various aspects of network design, network management, performance analysis and network integrity. Until June 1996, Dr. Pack was Executive Director of Network Integrity Planning at Bell Communications Research (Bellcore), where he successfully managed software, modeling and consulting business of $10-70M. He has played a prominent role in developing innovative methods for traffic engineering, network planning, dynamic routing, forecasting, data analysis, and demand modeling. He and his staff developed many of the procedures that have been standardized for use in engineering telecommunications networks, both domestically and internationally. From 1986 until 1991, he provided technical support for the LECs at the ICCF's Availability Workshops and the T1Q1 performance standards meetings that developed existing trunk blockage measurements and reports for InterLATA access.

Dr. Pack has more than 40 publications, including an award winning paper on statistical sampling and a book (co-edited with Dr. J. W. Cohen) on ATM performance. He is currently a Visiting Professor in Monmouth University's Computer Science Department, where he is teaching, doing research and providing consulting services to the Navy and several telecommunications companies. Dr. Pack is a Senior Member of the IEEE, a member of the Operations Research Society of America and the American Mathematical Association and is one of only two United States members of the International Advisory Council of the prestigious International Teletraffic Congress (ITC). In 1995, he was named a Bellcore Fellow. He is an editor of the international journal, "Advances in Performance Analysis".

**J. James Gordon,** B.Sc., Ph.D.
President, CAPE Consulting, Rumson, New Jersey

James Gordon holds a Bachelor of Science degree from the University of Queensland, Australia, and a Ph.D. from the University of Tasmania, Australia. Since completing his Ph.D., he has held positions as Research Associate in the Teletraffic Research Center, University of Adelaide, Australia, and Assistant Professor of Teletraffic Science, Bond University, Australia. In 1991 he joined the Network Integrity Planning organization at Bell Communications Research (Bellcore), New Jersey, where he worked on a variety of projects relating to the performance and engineering of telecommunications equipment. This work included technical auditing of Signaling System 7 (SS7) equipment, SS7 network outage analysis for RBOCs, technical lead of a project to analyze and evaluate SCP and intelligent peripheral products, analysis of next generation multi-processor switch architectures, and personal communications service (PCS) network architectures.

In 1997, Dr. Gordon started CAPE Consulting with two colleagues, with the goal of providing specialist capacity and performing engineering consulting services to the telecommunications industry. Through CAPE Consulting he continues to do engineering work for Bellcore and other clients. Notably, over the past two years, he has acted as a consultant to Bellcore and its clients on the impact of dialup internet traffic on the Public Switched Telephone Network. He has co-authored two Bellcore white papers and a magazine article which have been widely referenced in the technical and general press as authorities on this subject. He was instrumental in exploring the impact of internet traffic on LEC trunk engineering, and has proposed and led consulting projects for LECs in this area. He has published more than 20 journal articles, conference papers and magazine articles, and is a member of the IEEE.

**Albert A. Fredericks, Ph.D.**
Professor, Monmouth University, West Long Branch, New Jersey

Dr. Fredericks is Professor of Computer Science at Monmouth University, West Long Branch, New Jersey. From 1982 to 1990 he headed the Performance Analysis Department at AT&T Bell Labs. Responsible for directing and managing research, exploratory development, consulting activities and educational services to provide AT&T with the theory, methods, tools and support it needed to ensure that its products and services met performance requirements in a cost effective manner. Dr. Fredericks' organization was responsible for supporting the design, development, production and operation of most major AT&T products and services. Exploratory software development work included a variety of systems integrating expert system technology with quantitative analysis tools, particularly for resource management. (One result was the commercial product Q+, AT&T's Performance Analysis Workstation.)

Dr. Fredericks currently serves as the Graduate Program Director for Computer Science and as Director of Monmouth's Simulation and Modeling Laboratory. He is involved in a variety of consulting projects both as a member of Monmouth's Center for Technology Development and Transfer and in association with Quantitative Insights Systems and Services, Inc. His consulting work includes performance modeling, analysis, engineering and management of government (C3I) and commercial communications and information systems and the development and delivery of customized training and education courses in these areas. Dr. Fredericks is also responsible for design and prototype development of an open, object-oriented environment for modeling, analysis, simulation and engineering of computer / communications (and other) systems.

# DEFINITIONS OF TRUNK PERFORMANCE: TRUNK BLOCKAGE VERSUS CALL COMPLETION

## 1. CONCEPTS OF PERFORMANCE AND PARITY

The Commission's NPRM 98-72 addresses the issue of how parity should be defined when comparing performance of CLEC trunk groups and ILEC (incumbent LEC) trunk groups. In order to compare the performance of two trunk groups, one must first identify a measure or metric of performance. There exists a body of established practices, mathematical approaches and network measurements relating to trunk engineering and performance measurement. This body of practices was developed primarily to support methods and tools for the internal engineering of ILEC trunk groups, and to facilitate identification and resolution of performance problems associated with IXC traffic flowing over common (LEC) trunk groups. It is likely that this body of knowledge will prove helpful in defining parity between CLEC and ILEC networks. However, to the extent that CLECs are exploiting new types of network interconnection arrangements, e.g., those enabled by the Telecommunications Act of 1996, the authors believe that traditional measurement approaches may need to be extended and re-validated in order to arrive at a satisfactory framework for assessing CLEC / ILEC trunking parity. In that regard, this paper discusses the novel use of "call completion" statistics for assessing trunk network performance.

Independent of any particular performance metric, we believe that it is useful to propose some key principles for a **qualitative** definition of CLEC / ILEC trunk performance parity. The authors believe that an ILEC may **fail** to provide parity of performance if it can be shown that all three of the following basic factors are true:

1. there is a statistically significant difference in performance,

2. there is a persistent difference in performance (over some appropriately defined interval),

3. and the problem is generally under ILEC control.

The three elements in the above-proposed definition are important for the following reasons. Telecommunications traffic contains an inherent property of volatility or statistical variability. Phone calls do not arrive at pre-determined times. Instead, they arrive according to some random process, that may be characterized via its statistical properties. Similarly, traffic exhibits volatility from hour to hour and day to day, some of which is driven by external events such as the weather, promotional campaigns, political events, catastrophes and so on. For these reasons, trunk performance must be measured via appropriate statistics. It follows that in order to say that two trunk groups are experiencing different levels of service, one must ensure that the measured difference in performance (e.g., between two 20-day averages) is statistically significant, i.e., that it is not simply an artifact of the inherent volatility of the traffic.

Persistence is an intuitively reasonable property of a parity measure. One does not want to "falsely" determine that parity is not provided based, for example, on a one day glitch in performance data. It is reasonable to determine that parity has not been provided only if a performance discrepancy exists over a longer period of time. In practice, one can filter out transient problems, and detect "real" differences in service, by averaging performance data over an appropriate reporting interval. Traditionally, this has been a 20-day period (business month) as defined in documents such as Bellcore's SR-TAP-000191.

Finally, it is important to recognize that an ILEC requires certain input data and advance notice from all CLECs (and IXCs) in order to satisfactorily engineer its trunk groups; this includes forecasts of near-term traffic demands and network rearrangements. Trunks cannot be provisioned instantaneously. Instead, one must plan ahead based on forecast data, and allow for a lead time in equipment installation. If the ILEC relies on a CLEC for certain engineering information (e.g., forecast data), and the quality of this data is poor, then this can lead to performance problems. In a similar manner, CLEC network failures, equipment outages, translations error, routing errors and other similar problems can temporarily create performance problems in the ILEC network. Under either of these general categories of circumstances, it would be unreasonable to hold the ILEC responsible for the poor performance. For this reason, it is proposed that any determination of parity failure should be conditional on the factors contributing to the failure being generally within the control of the ILEC.

We believe that the above three elements – significance, persistence and ILEC control – are intuitively reasonable. The difficulty in arriving at a concrete definition of parity is to make these concepts precise: How long does a problem have to occur before it is deemed to be persistent? What are reasonable thresholds to be used in determining statistical significance? How can these thresholds be designed to reflect the inherent traffic volatility due to peakedness and day-to-day variation? And, at what point does one decide that CLEC data is of inferior quality, and that the ILEC is no longer 'responsible' for the performance problem. This document will not answer these questions, since complete answers will depend on detailed data analysis. However, it will outline a framework that the authors believe can be used to successfully answer the above questions, and arrive at a workable process for parity determination.

## 2. CUSTOMER QOS VERSUS NETWORK QOS

At a technical level, it is important to distinguish between **customer** Quality of Service (QOS) and **network** QOS. Customer QOS is the level of service perceived by the customer, on an 'end to end' basis, without regard to the details of the underlying network transport. Network QOS is a more direct measure of the quality of service from a network perspective. For example, the percent calls blocked on a given trunk group is a measure of network QOS. The customer, on the other hand, may not be concerned with the blocking probability on particular trunk groups, as long as his or her call completes via alternative routes. Customer QOS and network QOS are different. In describing the performance of telecommunications networks, it is intuitively natural to start with measures of customer QOS. For example, it may be desirable to have objectives such as:

- block no more than 1% of calls going from Chicago to Detroit, or
- block no more than 1% of calls carried by any CLEC or ILEC, or
- block no more than .5% of any calls routing via a tandem to a CLEC or an IXC.

However, in the context of circuit-switched networks with hierarchical traffic routing, it is very difficult, and perhaps impossible from a practical perspective, to ***engineer and operate*** LEC networks to meet such a ***customer QOS*** objective. The difficulties relate to the highly structured routing arrangements, ***for most traffic***, wherein a fixed sequence of **trunk groups** are searched to find an available path from point A to point B in the network.